

Precision matrix expansion – efficient use of numerical simulations in estimating errors on cosmological parameters

Oliver Friedrich^{1,2★} and Tim Eifler^{3,4}

¹Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr 1, D-81679 München, Germany

²Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany

³Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

⁴Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA

Accepted 2017 September 29. Received 2017 September 6; in original form 2017 March 21

ABSTRACT

Computing the inverse covariance matrix (or precision matrix) of large data vectors is crucial in weak lensing (and multiprobe) analyses of the large-scale structure of the Universe. Analytically computed covariances are noise-free and hence straightforward to invert; however, the model approximations might be insufficient for the statistical precision of future cosmological data. Estimating covariances from numerical simulations improves on these approximations, but the sample covariance estimator is inherently noisy, which introduces uncertainties in the error bars on cosmological parameters and also additional scatter in their best-fitting values. For future surveys, reducing both effects to an acceptable level requires an unfeasibly large number of simulations. In this paper we describe a way to expand the precision matrix around a covariance model and show how to estimate the leading order terms of this expansion from simulations. This is especially powerful if the covariance matrix is the sum of two contributions, $\mathbf{C} = \mathbf{A} + \mathbf{B}$, where \mathbf{A} is well understood analytically and can be turned off in simulations (e.g. shape noise for cosmic shear) to yield a direct estimate of \mathbf{B} . We test our method in mock experiments resembling tomographic weak lensing data vectors from the Dark Energy Survey (DES) and the Large Synoptic Survey Telescope (LSST). For DES we find that 400 N -body simulations are sufficient to achieve negligible statistical uncertainties on parameter constraints. For LSST this is achieved with 2400 simulations. The standard covariance estimator would require $>10^5$ simulations to reach a similar precision. We extend our analysis to a DES multiprobe case finding a similar performance.

Key words: methods: statistical – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

Wide area surveys such as the currently running Dark Energy Survey (DES, Flaugher 2005) or the upcoming Large Synoptic Survey Telescope (LSST, Ivezić et al. 2009) will collect vast amounts of data about the large-scale structure on the Universe. In cosmological analyses this data can e.g. be compressed into measurements of two-point correlation functions of galaxy clustering or cosmic shear. In a redshift-tomographic analysis this will easily accumulate to data vectors with several hundreds of data points. Testing cosmological models from a measurement of such a large data vector requires precise knowledge of the covariance matrix of the noise in this data vector and especially of the inverse covariance, which is also called the precision matrix. To obtain good estimates of these matrices,

survey collaborations have in the past e.g. taken the following route (Heymans et al. 2013; Kilbinger et al. 2013; Becker et al. 2016): they ran a set of high-precision numerical simulations, that in the limit of infinite realizations would in principle allow for a calculation of the true underlying covariance matrix of their observable. Given that numerical simulations are expensive, they however estimated the covariance and precision matrix from only a limited amount of realizations, leaving them with possibly significant uncertainties of their exact error budget.

There has been extensive research on the impact of errors associated with covariance estimation on the constraints derived on cosmological parameters. Hartlap, Simon & Schneider (2007) discussed the fact that the inverse of an unbiased covariance estimator is not an unbiased estimator for the inverse covariance matrix (*the precision matrix*). They also described a way to correct for this when assuming that the covariance estimate follows a Wishart distribution (see also Kaufman 1967 and

* E-mail: oliverf@usm.uni-muenchen.de

Anderson 2003). The noise properties of this corrected precision matrix estimator and its impact on the constraints derived on cosmological parameters was e.g. investigated by Taylor, Joachimi & Kitching (2013), Dodelson & Schneider (2013), Taylor & Joachimi (2014).

Sellentin & Heavens (2016, hereafter SH16a) have presented a different approach: given a covariance estimate they marginalize over the posterior distribution of the precision matrix to compute the likelihood in parameter space. Assuming that the covariance estimate follows a Wishart distribution they have derived a simple, closed-form expression for the resulting likelihood function. In Sellentin & Heavens (2017) they have extended these results to derive the information loss in parameter space due to noisy covariance estimates. A fully non-Gaussian treatment of the effects discussed in Dodelson & Schneider (2013, hereafter DS13) is however still missing.

An important result of the above-mentioned works is the following: even if a set of simulations is large enough to give a precise estimate of the covariance, the process of inverting this covariance estimate amplifies the noise of the estimation in a way that can still significantly impact the constraining power of a cosmological analysis, i.e. the uncertainty of the derived parameter errors can still be comparable to the errors themselves (see e.g. equations 27 and 28 of Dodelson & Schneider 2013). To overcome this problem it is important to note that often major contributions to the covariance matrix are analytically well understood. Estimating these contributions with the help of simulated data is hence a significant waste of computational resources and incorporating prior knowledge about the covariance into the estimate can potentially decrease uncertainties on the error budget of a cosmological analysis.

Prior knowledge on the sparsity of the covariance matrix and the precision matrix was used by Paz & Sánchez (2015) and Padmanabhan et al. (2016) to improve estimates of the precision matrix from few simulations. Pope & Szapudi (2008) investigated shrinkage estimators of the covariance, i.e. a mixing of estimated and modelled covariance matrices. This however raises the task of finding an equivalent to the Kaufman–Hartlap correction for such a mixture of estimated and analytic matrices. More recently, Joachimi (2017) describes a non-linear extension of that estimator which combines covariance estimates from two sets of independent data vector realizations and hence does not require a covariance model.

In this paper we introduce a way of incorporating prior knowledge about the covariance directly into an estimate of the inverse covariance matrix, i.e. the precision matrix. We describe a way to expand the precision matrix around a covariance model as a power series in the deviation between model and true covariance. Assuming a Wishart realization for the true covariance (e.g. an estimate from N -body simulations) and using the results on invariant moments of the Wishart distribution by Letac & Massam (2004) we derive an unbiased estimator for the up to second-order expansion of the true precision matrix. This becomes especially powerful if parts of the covariance matrix that are well understood analytically can be turned off in simulations in order to yield a direct estimate of the remaining covariance parts. In Section 3 we recap the main problems of estimating parameter constraints from noisy covariance estimates and present our method of ‘precision matrix expansion’ (PME). In Section 4 we perform numerical experiments that mimic data from the DES and the LSST likelihood analyses to test the performance of our idea. Section 5 concludes with a discussion of our results.

2 PARAMETER CONSTRAINTS FROM NOISY COVARIANCE ESTIMATES

We begin by outlining the main task of this paper. Let $\hat{\xi}$ be a vector of N_d data points measured from observational data and let $\xi[\pi]$ be a model for this data vector that depends on a vector of N_p parameters π . If \mathbf{C} is the covariance matrix of $\hat{\xi}$ then a standard way to constrain the parameters π is to assign a posterior distribution $p(\pi|\hat{\xi})$ to them as

$$p(\pi|\hat{\xi}) \sim \exp\left(-\frac{1}{2}\chi^2[\pi|\hat{\xi}, \mathbf{C}]\right) p(\pi) \quad (1)$$

with

$$\chi^2[\pi|\hat{\xi}, \mathbf{C}] = (\hat{\xi} - \xi[\pi])^T \mathbf{C}^{-1} (\hat{\xi} - \xi[\pi]) \quad (2)$$

and $p(\pi)$ being a prior density incorporating a priori knowledge or assumptions on π . These expressions in fact ignore that \mathbf{C} also can be dependent on π . We will do this throughout this paper and refer the reader to Eifler, Schneider & Hartlap (2009) who investigated the impact of cosmology-dependent covariance matrices on cosmic shear likelihood analyses. Another assumption that goes into equation (1) is that the measured data vector $\hat{\xi}$ is drawn from a multivariate Gaussian distribution. In wide area surveys this is justified in the limit where one can consider the survey to consist of many independent sub-regions, such that the measurements in those regions add up to a Gaussian data vector by means of the central limit theorem.

If the covariance matrix \mathbf{C} is not exactly known, it can e.g. be estimated from N -body simulations. If $\hat{\xi}_i$, $i = 1 \dots N_s$, are a number of independent measurements of ξ in simulations then an unbiased estimate of \mathbf{C} is given by

$$\hat{\mathbf{C}} := \frac{1}{\nu} \sum_{i=1}^{N_s} (\hat{\xi}_i - \bar{\xi})(\hat{\xi}_i - \bar{\xi})^T, \quad (3)$$

where $\nu = N_s - 1$ and $\bar{\xi}$ is the sample mean of the $\hat{\xi}_i$. We will assume $\hat{\mathbf{C}}$ to have a Wishart distribution with ν degrees of freedom which follows from our assumption that $\hat{\xi}$ and the $\hat{\xi}_i$ are Gaussian distributed (cf. Taylor et al. 2013). Also, we will assume that $\hat{\mathbf{C}}$ is an unbiased estimator for the covariance matrix of actual data, i.e. if $\hat{\mathbf{C}}$ is indeed an estimate from N -body simulations, then we will assume these simulations to well resemble the error contributions present in actual data.

To compute the likelihood in equation (1) we need to know the precision matrix, i.e. is the inverse covariance matrix $\Psi = \mathbf{C}^{-1}$. According to Kaufman (1967, see also Hartlap et al. 2007; Taylor et al. 2013) an unbiased estimator for Ψ can be constructed from $\hat{\mathbf{C}}$ as

$$\hat{\Psi} = \frac{\nu - N_d - 1}{\nu} \hat{\mathbf{C}}^{-1} \quad (4)$$

and we will call the factor of $(\nu - N_d - 1)/\nu$ the Kaufman–Hartlap correction.

Given a measurement $\hat{\xi}$ of the data vector one can derive the posterior density of the model parameters $p(\pi|\hat{\xi})$ by means of equations (1) and (2). A noisy precision matrix estimate influences this inference in two ways:

- (i) it adds noise to the width of likelihood contours derived from inserting the precision matrix estimate into the figure of merit χ^2 (equation 2).
- (ii) it adds noise to the location of likelihood contours. Consider e.g. the maximum likelihood estimator for the parameters, $\hat{\pi}_{\text{ML}}$,

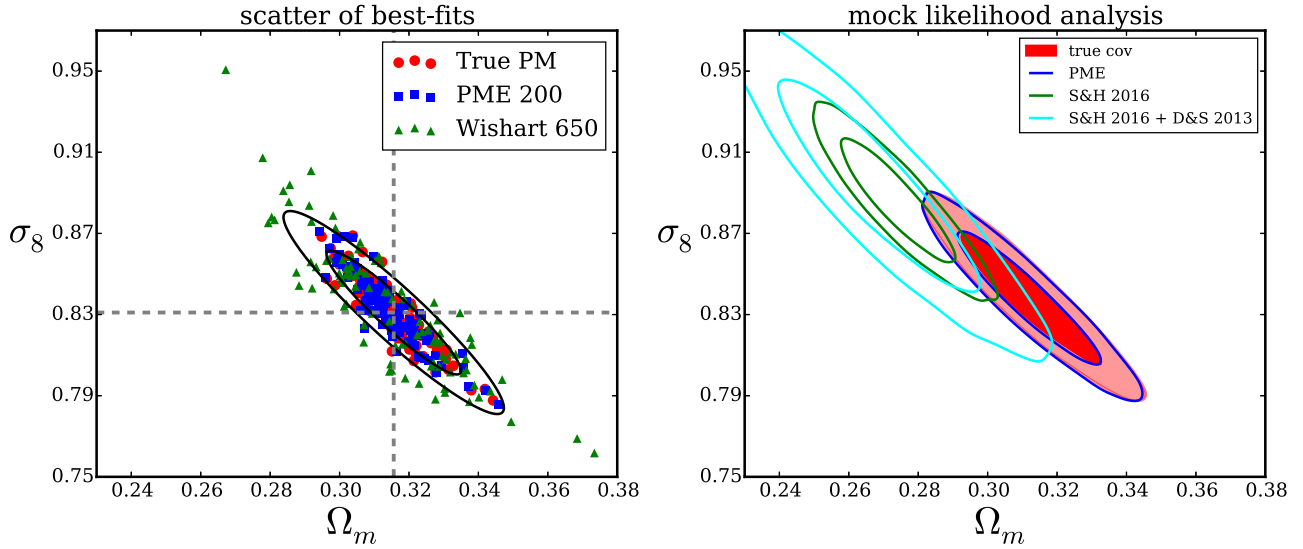


Figure 1. Left: Best-fitting parameter pairs (Ω_m, σ_8) obtained from random realizations of a DES-like weak lensing data vector with 450 data points when using different approaches to compute the precision matrix. The red points assume that the true covariance matrix is known while for the green points we draw a Wishart realization of the covariance ($N_s = 450 + 200 = 650$ simulations) for each data vector. The blue points are obtained with the method of precision matrix expansion (and allowing only 200 simulations to estimate the expansion). The black contours display the 1σ and 2σ Fisher contours derived from our fiducial covariance. Right: For one of the random realizations we perform a complete likelihood analysis and show the 1σ and 2σ contours in the Ω_m – σ_8 plane after marginalizing over w_0 and w_a (see Section 4 for details). The contours obtained from the Wishart realization of the covariance are clearly offset from those obtained from the true covariance matrix of our mock data vectors. We tried to account for this by expanding the likelihood around its point of maximum likelihood with the factor derived by DS13 (see main text for a more detailed description). This leads to a decreased constraining power of our mock survey. The use of PME manages to significantly decrease this contour offset.

which minimizes the χ^2 figure-of-merit,

$$\begin{aligned} & \left(\hat{\xi} - \xi[\hat{\pi}_{\text{ML}}] \right)^T \hat{\Psi} \left(\hat{\xi} - \xi[\hat{\pi}_{\text{ML}}] \right) \\ &= \min \left\{ \left(\hat{\xi} - \xi[\pi] \right)^T \hat{\Psi} \left(\hat{\xi} - \xi[\pi] \right) \right\}. \end{aligned} \quad (5)$$

When using a noisy precision matrix the uncertainties of $\hat{\pi}_{\text{ML}}$ have contributions from both the noise in $\hat{\xi}$ and the noise in $\hat{\Psi}$.

The astro-statistics literature has so far focused on the first effect, i.e. on the uncertainties on contour width due to noise in the estimate $\hat{\Psi}$ (Taylor et al. 2013; Taylor & Joachimi 2014; Sellentin & Heavens 2016, 2017). Sellentin & Heavens (2017) provide the most complete demonstration that $\hat{\Psi}$ yields a good estimate of the width of the posterior contours as long as $N_s - N_d \gg N_p$.

The more critical effect however is the additional noise of $\hat{\pi}_{\text{ML}}$. DS13 (also see Appendix A) showed that the uncertainty on the position of likelihood contours from noise in $\hat{\Psi}$ is only negligible if $N_s - N_d \gg N_d - N_p$ which is a much more demanding criterion for current cosmological data vectors. We demonstrate this in the left-hand panel of Fig. 1, where we show 100 randomly drawn realizations of a DES-like weak lensing data vector with $N_d = 450$ and a halo model covariance matrix (see Section 4 for further details). For each of the 100 data vectors we have also generated Wishart realizations of our covariance matrix corresponding to an estimate from $N_s = 650$ simulations. Using either the true covariance of our mock data vectors or the estimated one, we then determine the best-fitting parameters Ω_m and σ_8 (after marginalizing over equation-of-state parameters of dark energy, w_0 and w_a). The best fits obtained from a noisy covariance (green points) clearly display a much larger scatter than those obtained from the true covariance (red points). Also shown are the best fits obtained by PME (blue points) which we are going to introduce in the next section. Here we assumed that only

$N_s = 200$ simulations are available to estimate the PME, which gives best-fitting values that are significantly closer to the ones obtained when knowing the true covariance matrix.

When reconstructing $p(\pi|\hat{\xi})$ (e.g. from a Monte Carlo-Markov chain) this can lead to significant offsets between likelihood contours inferred from a covariance estimate and the likelihood contours that would be inferred if the true covariance was known – even if the overall width of the likelihood contours is captured well by the covariance estimate. We demonstrate this in the right-hand panel of Fig. 1. DS13 have derived a factor (see Appendix A) by which parameter contours obtained from a Wishart realization of the covariance should be expanded in order to account for this additional scatter. However, their derivation relies on the assumption of a Gaussian parameter likelihood and is only applicable to the extent that a Fisher analysis is accurate.

To nevertheless demonstrate the effect of the DS13 correction, we propose the following combination of SH16a and DS13: based on our MCMC we determine the point of maximum likelihood in our full (in this case four-dimensional) parameter space. Then we expand the distance of each parameter point in our chain to the maximum likelihood point by the square root of the correction factor presented in Appendix A. This procedure reduces exactly to the correction derived by DS13 in the case of a Gaussian parameter likelihood. Also it leaves the point of maximum likelihood in the reconstructed likelihood function unchanged. We implement this idea for the cyan contours in Fig. 1 and show that this brings the contours derived from a standard covariance estimate into consistency with those derived from the true covariance of our mock data vectors.

Downsides of this approach are a large increase of the uncertainties on cosmological parameters and the fact that one still needs at least as many realizations as data points in the data vector to even derive a precision matrix estimate. We now want to introduce an

alternative method to estimate the precision matrix which is able to drastically decrease the offset of contours seen for the standard precision matrix estimator.

3 PRECISION MATRIX EXPANSION

Let us split the covariance matrix \mathbf{C} into two contributions

$$\mathbf{C} = \mathbf{A} + \mathbf{B}, \quad (6)$$

where for matrix \mathbf{A} we have an accurate model (e.g. the shape-noise contributions to the covariance of cosmic shear correlation functions) and for \mathbf{B} we have a model \mathbf{B}_m which we know to be imperfect. We want to include this prior knowledge of the covariance matrix when estimating the precision matrix. Starting from

$$\mathbf{C} = \mathbf{M} + (\mathbf{B} - \mathbf{B}_m), \quad (7)$$

where $\mathbf{M} = \mathbf{A} + \mathbf{B}_m$ is our model for the complete covariance matrix, we rewrite

$$\mathbf{C} = (\mathbb{1} + \mathbf{X}) \mathbf{M}, \quad (8)$$

where $\mathbb{1}$ is the identity matrix and we have defined

$$\mathbf{X} := (\mathbf{B} - \mathbf{B}_m) \mathbf{M}^{-1}. \quad (9)$$

The precision matrix $\Psi = \mathbf{C}^{-1}$ can then be expressed as the following power series in \mathbf{X} :

$$\begin{aligned} \Psi &= \mathbf{M}^{-1} \left(\sum_{k=0}^{\infty} (-1)^k \mathbf{X}^k \right) \\ &= \mathbf{M}^{-1} (\mathbb{1} - \mathbf{X} + \mathbf{X}^2 + \mathcal{O}[\mathbf{X}^3]). \end{aligned} \quad (10)$$

We will call this series the PME. In Appendix C we show that it converges under a wide range of conditions. There we also demonstrate that the series yields at each order a symmetric approximation of Ψ and that at second order it is always positive definite (at each order if the series converges).

3.1 Estimating the expansion of Ψ

Suppose we have an estimate $\hat{\mathbf{B}}$ of the matrix \mathbf{B} from a number of N -body simulations. This especially assumes that all covariance contributions included in \mathbf{A} can be turned off in the simulations (i.e. for cosmic shear covariances \mathbf{A} could consist of shape-noise contributions which can be set to zero in simulations). We want to use $\hat{\mathbf{B}}$ to construct unbiased estimators for the first-order and second-order term of the series in equation (10).

Our assumptions state that $\hat{\mathbf{B}}$ is drawn from a Wishart distribution with expectation value \mathbf{B} . In this case also $\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}$ is Wishart distributed but with the expectation value $\mathbf{M}^{-1}\mathbf{B}\mathbf{M}^{-1}$. Hence an unbiased estimator for the first-order PME is given by

$$\hat{\Psi}_{1st} = \mathbf{M}^{-1} - \mathbf{M}^{-1} (\hat{\mathbf{B}} - \mathbf{B}_m) \mathbf{M}^{-1}. \quad (11)$$

Note that this does not involve the inversion of an estimated matrix. According to Taylor et al. (2013) the standard deviation of diagonal elements of an inverse-Wishart distributed matrix is proportional to $1/\sqrt{N_s - N_d - 4}$ while for Wishart distributed matrices it is only proportional to $1/\sqrt{N_s - 1}$. Hence, avoiding the occurrence of an inverted matrix estimate greatly reduces the estimation noise.

The second-order term involves squares of Wishart matrices. Using the results of Letac & Massam (2004) on invariant moments of the Wishart distribution (cf. Appendix B) it is still possible to

Table 1. Number of tomographic bins, total number of data points and type of data vector for the different setups used to test the performance of precision matrix expansion.

Setup	Survey	Lens bins	Source bins	N_{data}	Data
Ia	DES	0	5	450	Real space
Ib	DES	3	5	630	Real space
II	LSST	0	10	2200	Fourier

construct an unbiased estimator for the second-order PME as

$$\begin{aligned} \hat{\Psi}_{2nd} &= \mathbf{M}^{-1} + \mathbf{M}^{-1} \mathbf{B}_m \mathbf{M}^{-1} \mathbf{B}_m \mathbf{M}^{-1} \\ &\quad - \mathbf{M}^{-1} (\hat{\mathbf{B}} - \mathbf{B}_m) \mathbf{M}^{-1} \\ &\quad - \mathbf{M}^{-1} \hat{\mathbf{B}} \mathbf{M}^{-1} \mathbf{B}_m \mathbf{M}^{-1} \\ &\quad - \mathbf{M}^{-1} \mathbf{B}_m \mathbf{M}^{-1} \hat{\mathbf{B}} \mathbf{M}^{-1} \\ &\quad + \mathbf{M}^{-1} \frac{v^2 \hat{\mathbf{B}} \mathbf{M}^{-1} \hat{\mathbf{B}} - v \hat{\mathbf{B}} \text{tr}(\mathbf{M}^{-1} \hat{\mathbf{B}})}{v^2 + v - 2} \mathbf{M}^{-1}. \end{aligned} \quad (12)$$

The estimator in equation (12) is the key result of our paper. It has two advantages over the Anderson–Hartlap corrected standard estimator. First, it only requires matrix multiplications. As a consequence, it can even be used if $N_s \leq N_d$. Secondly, it only needs an estimate of \mathbf{B} instead of the whole covariance \mathbf{C} , i.e. it allows to incorporate a priori knowledge on the covariance in the form of \mathbf{M} (and \mathbf{A}).

In the next section we demonstrate that this significantly eases the requirement of $N_s - N_d \gg N_d - N_p$. Hence, in a likelihood analysis the noise in $\hat{\Psi}_{2nd}$ becomes negligible for a much smaller number of N -body simulations than required by the standard precision matrix estimator. In Appendix C we also show that the bias in parameter constraints which arises from cutting the power series in equation (10) after a finite number of terms is negligible even for very strong deviations of our covariance model \mathbf{M} from the N -body covariance \mathbf{C} .

4 EXAMPLES: PARAMETER ERRORS FOR LSST WEAK LENSING AND DES WEAK LENSING AND MULTIPROBE ANALYSES

We investigate the performance of our method in the context of ongoing and future surveys using DES and LSST as specific examples. These surveys differ in terms of survey area, galaxy number density and redshift distribution, and have different demands on the precision matrix. For DES we consider summary statistics in real space, i.e. auto- and cross-correlation functions of galaxy shear and position; for LSST we consider the corresponding Fourier quantities of a shear–shear only data vector. A summary of the scenarios considered is given in Table 1 and a more detailed description of the considered data vectors is given in Appendix D.

In order to test the performance of PME we set up mock experiments where we assume the true covariance matrix of each survey to be the analytic halo-model covariance described in Krause & Eifler (2017). This model divides the covariance into three contributions: a noise-only part that consists of shape- and shot-noise contributions, \mathbf{C}^{nn} , a contribution from the cosmic variance of the signal, $\mathbf{C}^{\text{ss,halo}}$, and a mixed term including noise and signal contributions, \mathbf{C}^{sn} . For shear–shear only covariances we set

$$\mathbf{A} = \mathbf{C}^{\text{nn}} + \mathbf{C}^{\text{sn}} \quad (13)$$

and

$$\mathbf{B} = \mathbf{C}^{\text{ss,halo}}. \quad (14)$$

The shape-noise contributions to the covariance can be modelled reliably since the ellipticity dispersion can be measured from the data itself and since the mixed term \mathbf{C}^{sn} involves only the modelling of two-point statistics of the shear field. The \mathbf{B} term comprises the more complex four-point statistics of the shear field, which can be estimated from simulations by turning off shape noise. This is more complicated for galaxy clustering where shot noise is included in the covariance matrix (cf. Section 4.1.2).

In order to simulate a situation where our covariance model $\mathbf{M} = \mathbf{A} + \mathbf{B}_m$ deviates from the true covariance we degrade it as

$$\mathbf{B}_m = \alpha \mathbf{C}^{\text{ss,Gauss}} + \beta (\mathbf{C}^{\text{ss,halo}} - \mathbf{C}^{\text{ss,Gauss}}), \quad (15)$$

where $\mathbf{C}^{\text{ss,Gauss}}$ contains only the parts of the cosmic variance that are also present in a Gaussian covariance model. Hence, we allow the Gaussian and non-Gaussian cosmic variance parts to be over- or underestimated by a constant multiplicative factor. If not stated differently in this section we will use $\alpha = 1.0$ and $\beta = 0.5$. In Appendix C we explore a wider range of rescalings and also consider more complex deformations of our fiducial covariance to show that the PME remains robust under more complicated deviations of the model covariance \mathbf{M} from the true covariance matrix. All simulated likelihood analyses in this paper are computed using the `COSMOLIKE` cosmology package (Eifler et al. 2014; Krause & Eifler 2017).

4.1 Performance for DES weak lensing data vector

We now carry out mock likelihood analyses for DES and LSST weak lensing data vectors, varying the parameters Ω_m , σ_8 , w_0 and w_a . Our fiducial values for these parameters are

$$(\Omega_m, \sigma_8, w_0, w_a) = (0.3156, 0.831, -1, 0). \quad (16)$$

We start by drawing random Gaussian realizations of our fiducial data vectors according to a covariance given by the halo model. For each realization we also draw new Wishart realizations $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ of cosmic variance and total covariance to compute the PME estimate $\hat{\Psi}_{2\text{nd}}$ and the standard estimator $\hat{\Psi}$. In practice, this is done by drawing additional realizations $\hat{\xi}_i$, $i = 1 \dots N_s$, of our fiducial data vector from a multivariate Gaussian distribution whose covariance is \mathbf{B} respectively \mathbf{C} . These realizations represent measurements from N -body simulations and inserting them into equation (3) generates the desired Wishart realizations $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ of the two matrices.

Using `COSMOLIKE` we then run likelihood chains to infer a posterior distribution for our parameters using either $\hat{\Psi}_{2\text{nd}}$, $\hat{\Psi}$ or the true precision matrix of our mock data vectors, \mathbf{C}^{-1} . When computing the likelihood from \mathbf{C}^{-1} and $\hat{\Psi}_{2\text{nd}}$ we simply use standard ansatz given in equation (1). When deriving contours from the Wishart realization $\hat{\mathbf{C}}$ we furthermore compute the parameter likelihood as

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{\xi}}) \sim \left[1 + \frac{(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}])^T \hat{\mathbf{C}}^{-1} (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}])}{N_s - 1} \right]^{-N_s/2}, \quad (17)$$

which SH16a have shown to be a more accurate than using the Kaufman–Hartlap correction and the standard Gaussian likelihood. We however found only small differences to using the standard likelihood ansatz, which is due to the fact that in all cases considered in this paper $N_s - N_d \gg N_p$.

In Figs 2 and 3 we see the resulting 1σ and 2σ contours in the Ω_m – σ_8 plane (after marginalizing over the other parameters) for three different random draws of data vector and Wishart matrices. For each realization of the DES data vector we assumed that $N_s = 200$ simulations are available to estimate the PME and

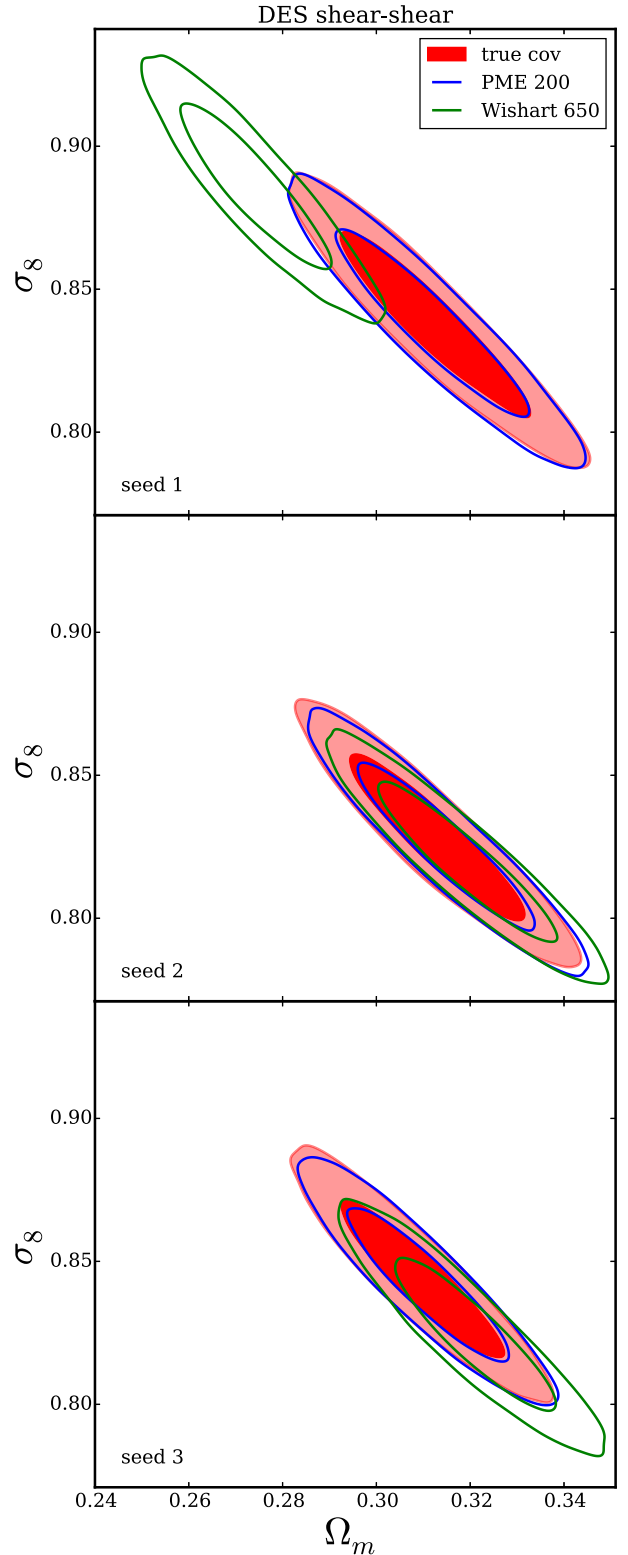


Figure 2. Contours in the Ω_m – σ_8 plane obtained from realizations of our DES-like weak lensing data vector after marginalizing over all other parameters. For each random seed also new Wishart realizations $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ of the matrices \mathbf{B} and \mathbf{C} were drawn in order to simulate new realizations of the second-order PME estimator and the standard precision matrix estimator. $N_s = 200$ simulations were assumed for the estimation of the PME while $N_s = N_d + 200 = 650$ simulations were assumed for the standard estimator.

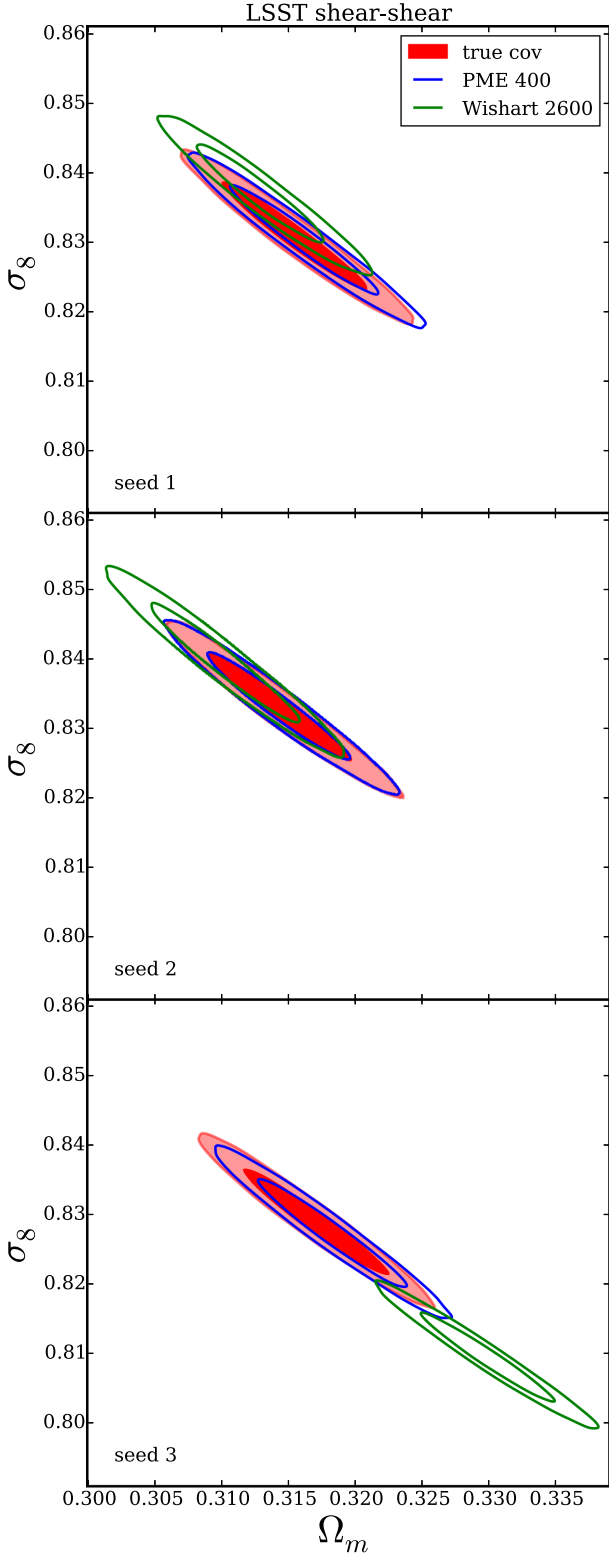


Figure 3. Same as Fig. 2 but for the LSST-like weak lensing data vector. $N_s = 400$ simulations were assumed for the estimation of the PME while $N_s = N_d + 400 = 2600$ simulations were assumed for the standard estimator.

$N_s = N_d + 200 = 650$ simulations for the standard estimator. For each realization of the LSST data vector we assumed $N_s = 400$ simulations for the PME and $N_s = N_d + 400 = 2600$ simulations for the standard estimator.

Even though we allowed in each case many more simulations for the standard estimator than for the PME, the PME yields contours that are significantly closer to the ones obtained from the true precision matrix of our experiment, \mathbf{C}^{-1} . In particular we find that deviations between the contours derived from $\hat{\Psi}_{2\text{nd}}$ and those derived from \mathbf{C}^{-1} are much smaller than the corresponding 1σ and 2σ uncertainties of the parameters.

Next we generalize the findings in Figs 2 and 3. We generate 1000 Wishart realizations of the matrices $\hat{\mathbf{C}}$ and $\hat{\mathbf{B}}$ for different assumptions on the number of available N -body simulations N_s . For each of the 1000 sets of matrices we also generate 10 realizations $\hat{\xi}$ of our fiducial data vector (i.e. overall 10 000 different realizations $\hat{\xi}$). Hence for each type of precision matrix estimate we perform overall 10 000 likelihood analyses. In each analysis we determine the best-fitting parameters $\hat{\pi}_{\text{ML}}$ and check whether our fiducial cosmology is outside the 68.3 per cent confidence contour around these parameters. In order to make this computationally feasible, we are now linearly approximating the calculations of COSMOLIKE around our fiducial cosmology π_0 , i.e. we use

$$\xi_{\text{simple}}[\pi] = \xi_{\text{exact}}[\pi_0] + \sum_{i=1}^{N_p} (\pi_i - \pi_{0,i}) \left. \frac{\partial \xi_{\text{exact}}}{\partial \pi_i} \right|_{\pi_0}. \quad (18)$$

Since our mock data vectors $\hat{\xi}$ are Gaussian, the best-fitting parameters $\hat{\pi}_{\text{ML}}$ of this linearized model will also have a multivariate Gaussian distribution where the inverse of the parameter covariance $\mathbf{F} = \mathbf{C}_{\text{param}}^{-1}$ is given by

$$F_{ij}[\mathbf{C}^{-1}] = \left. \frac{\partial \xi_{\text{exact}}^T}{\partial \pi_i} \right|_{\pi_0} \mathbf{C}^{-1} \left. \frac{\partial \xi_{\text{exact}}}{\partial \pi_j} \right|_{\pi_0}. \quad (19)$$

In this simplified situation $F_{ij}[\mathbf{C}^{-1}]$ is identical to the Fisher matrix and we will interchangeably call it Fisher matrix or inverse parameter covariance. We can also use the precision matrix estimates $\hat{\Psi}$ and $\hat{\Psi}_{2\text{nd}}$ to estimate the inverse parameter covariance as $F_{ij}[\hat{\Psi}]$ and $F_{ij}[\hat{\Psi}_{2\text{nd}}]$. This allows us to analytically determine the maximum likelihood parameters and the 68.3 per cent confidence contours that would be obtained from each precision matrix estimate and each random realization of our data vector, $\hat{\xi}$. Note that we only need these approximations to make our analysis computationally feasible. An additional benefit is though that the results of DS13 hold exactly in this simplified situation, allowing us to cross-check our results with their analytical findings.

We define $F_{>1\sigma}$ as the fraction of times that our fiducial cosmology is outside of the 68.3 per cent confidence contour around the best-fitting parameters and we use it as a metric for comparing the different precision matrix estimators. In Fig. 4 we show this fraction for all different types of precision matrices introduced before. The solid, dashed and dotted lines show the fractions achieved when using the noise-less matrices \mathbf{C}^{-1} , \mathbf{M}^{-1} and $\Psi_{2\text{nd}}$. Especially, the noise-less matrix $\Psi_{2\text{nd}}$ would be the PME estimator in the limit of infinitely many simulations and \mathbf{C}^{-1} would be the standard estimator in the same limit. The red and blue dots show the fraction achieved when using the noisy precision matrix estimates $\hat{\Psi}$ and $\hat{\Psi}_{2\text{nd}}$.

As expected, $F_{>1\sigma}$ is very close to 32 per cent when using the true covariance of our experiment, \mathbf{C} , in the likelihood analyses. For the deformed halo model covariance \mathbf{M} we assumed the two cases

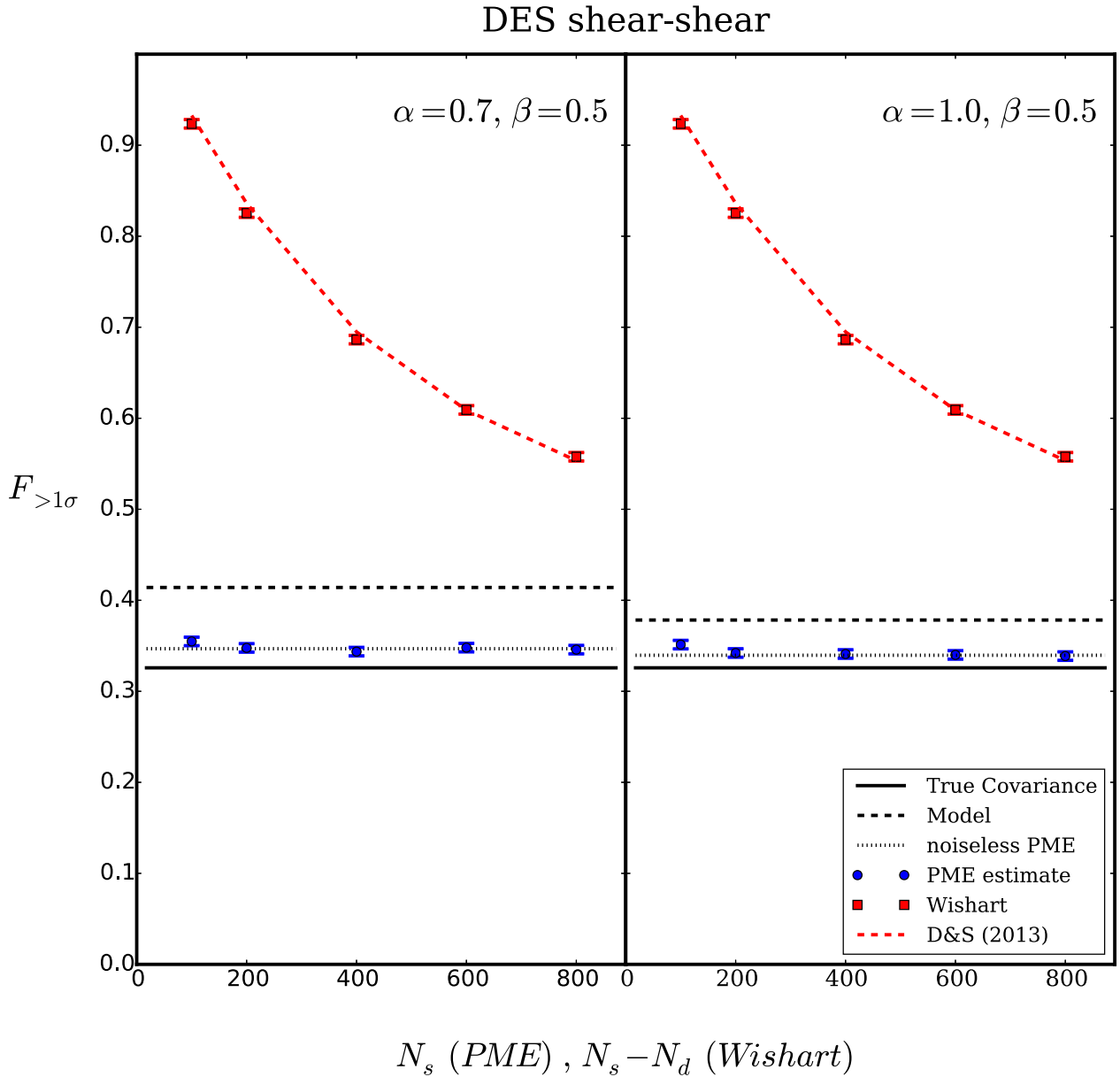


Figure 4. The figure compares $F_{>1\sigma}$, the number of times that our fiducial cosmology was considered outside the 68.3 per cent confidence contour in our simulated likelihood analyses when using different precision matrix estimates for computing the posterior parameter likelihood. In order to carry a sufficient number of mock analyses, we simplified our modelling of the data vector by linearly approximating the full computation around our fiducial cosmology. For the DES-like weak lensing data vector we varied the four parameters (Ω_m , σ_8 , w_0 , w_a).

$\alpha = 0.7$, $\beta = 0.5$ (left-hand panel) and $\alpha = 1.0$ and $\beta = 0.5$ (right-hand panel). For $\alpha = 0.7$ and $\beta = 0.5$ our fiducial cosmology is regarded as outside the 68.3 per cent contour in more than 40 per cent of the cases. For both choices of \mathbf{M} the noise-free PME significantly corrects that fraction towards the optimal value of ~ 32 per cent. Especially promising is that the PME estimate performs very similar to the noise-free PME. If 200 simulations are available to estimate the PME, it essentially converges to its best possible performance. And even if only 100 simulations are available to estimate the PME, its value of $F_{>1\sigma}$ comes closer to 32 per cent than when using \mathbf{M} to derive the contours.

When inferring the likelihood from the standard precision matrix estimator $F_{>1\sigma}$ is greater than 50 per cent even if we allow $N_s = N_d + 800$ simulations for the covariance estimation, which corresponds to 1250 simulations. This is due to the additional variance

of $\hat{\pi}_{\text{ML}}$ caused by the noise of the precision matrix (cf. equation 5). Using the results of DS13 we can derive predictions for this effect (cf. Appendix A). As can be seen from the red dashed lines in Fig. 4 these predictions agree well with what we find in our simulated likelihood analyses. Extrapolating the results of DS13 to higher values of N_s we can also estimate, how many simulations would be required for the standard precision matrix estimator in order to achieve the same value of $F_{>1\sigma}$ as the second-order PME. For the left-hand panel of Fig. 4 we find that it would take ~ 8000 simulations for the standard estimator to get as close to $F_{>1\sigma} = 32$ per cent as the PME with only 200 simulations. This statement however depends on the model covariance \mathbf{M} since it determines how well the PME has converged after its second order.

An \mathbf{M} -independent way of comparing standard estimator and PME estimator is to see how many simulations it takes each to

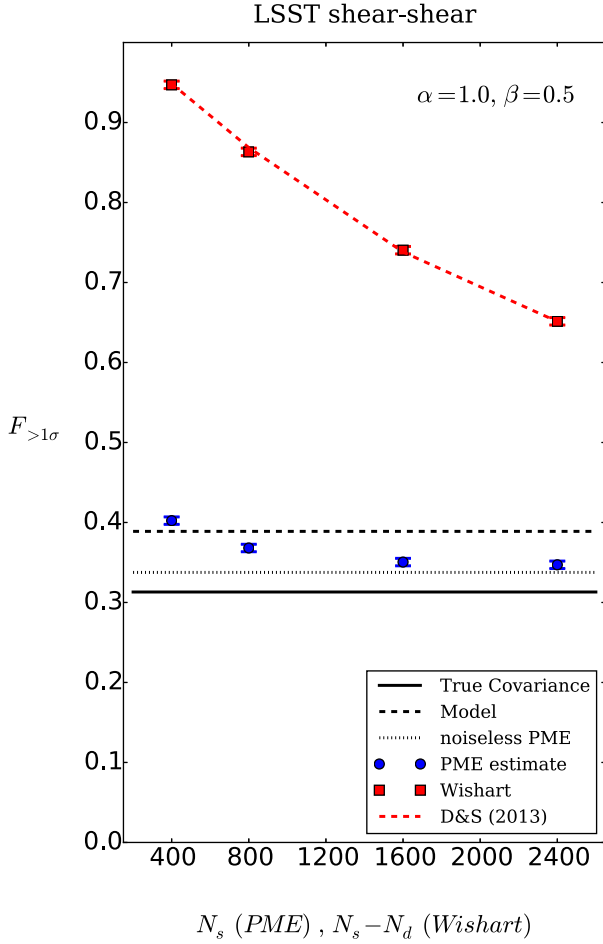


Figure 5. Same as Fig. 4 but for the LSST-like weak lensing data vector.

have $F_{>1\sigma}$ within 1 per cent of their best possible performance. It would take the standard estimator $\sim 24\,000$ simulations to be within 1 per cent of $F_{>1\sigma} = 32$ per cent. The PME estimator is well within 1 per cent of its best possible performance for only 200 simulations.

Note that with the results of DS13 one can in principle correct a likelihood analysis for the additional variance caused by the standard precision matrix estimator. This would result in a decreased constraining power of the analysis and it would hence be the main benefit of the PME to prevent this loss.

4.1.1 Larger covariance matrices: LSST weak lensing data vector

We repeat the above analysis for the LSST-like weak lensing data vector. Fig. 5 shows the fractions $F_{>1\sigma}$ obtained from PME and standard precision matrix estimator. The PME estimator now requires ~ 2400 simulations to be less than 1 per cent away from its best possible performance. As before, this statement does not include any additional biases between PME and true precision matrix that might arise from the biased model matrix \mathbf{M} used to carry out the matrix expansion. The standard precision matrix estimator would need $N_s > 115\,000$ simulations to be less than 1 per cent away from its best possible performance.

4.1.2 Defining \mathbf{A} and \mathbf{B} for multiprobe covariances

We now repeat the analysis of Fig. 4 for a DES-like multiprobe data vector. This vector includes contributions from galaxy clustering

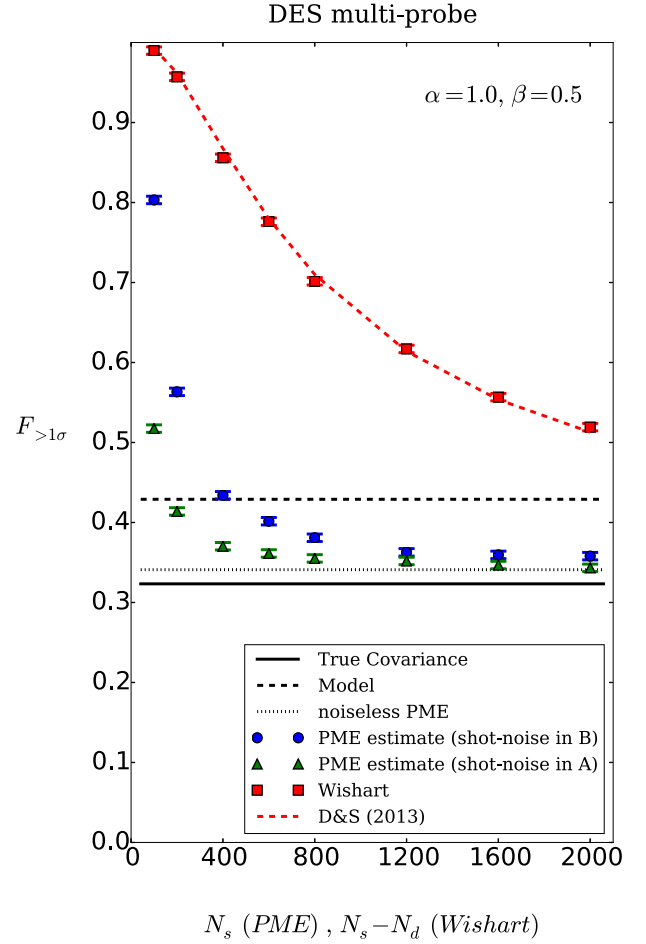


Figure 6. Same as Fig. 4 but for the DES multiprobe data vector. For this case the seven parameters (Ω_m , σ_8 , w_0 , w_a , b_1 , b_2 , b_3) were varied in each likelihood analysis. Green points assume that cosmic variance can be estimated from simulations without shot noise. This would significantly improve the performance of PME for low numbers of available simulations.

and galaxy–galaxy lensing, which introduces shot-noise terms to the covariance. These shot-noise contributions are in principle well understood theoretically and include, similar to the cosmic shear case, at most two-point statistics of the cosmic density field. Hence, one could absorb them into the matrix \mathbf{A} (cf. equation 6) and use N -body simulations only for the remaining part of the covariance – i.e. to define \mathbf{B} as only the cosmic variance. This is however difficult since most N -body simulations provide only simulated galaxy catalogues that are affected by shot noise themselves, which makes it impossible to independently estimate the cosmic variance. If however all shot-noise contributions are included in \mathbf{B} when defining and estimating the PME, then the estimator $\hat{\Psi}_{2nd}$ will have a higher variance in many of its elements. Hence, the additional scatter of best-fitting parameters due to a noisy precision matrix might not be negligible anymore.

In Fig. 6 we compare the fractions $F_{>1\sigma}$ obtained from different estimates of the precision matrix in our simulated likelihood analyses – this time for the DES multiprobe data vector. In each likelihood analysis we now vary seven parameters, since for each lens bin we include a galaxy bias parameter in our model. The fiducial bias values are

$$(b_1, b_2, b_3) = (1.35, 1.50, 1.65) \quad (20)$$

in order of increasing redshift. Fig. 6 shows the results obtained for each of the mentioned options of defining **B**. It is clear that the noisy PME approaches its best possible performance already for a smaller number of simulations if the cosmic variance can be estimated directly. In practice this would however require density maps in thin redshift slices for each simulation in order to measure the correlation functions of the projected density fields without shot noise.

Assuming one can directly measure the cosmic variance from simulations we again want to assess how many simulations are required for the standard precision matrix estimator and the PME estimator to be within 1 per cent of their best possible performance. Extrapolating the results of DS13 we find that it would take the standard estimator $\sim 44\,000$ simulations to be within 1 per cent of $F_{>1\sigma} = 32$ per cent. The PME estimator is within 1 per cent of its best possible performance for 1600 simulations. For $N_s = 2000$ the performance of the PME becomes almost solely restricted by the deviation between **M** and **C** in our mock experiment. However, below $N_s = 1600$ there seems to be significant additional scatter of the best-fitting parameters due to the noise of the PME estimate. We demonstrate this in Fig. 7 for $N_s = 400$. Regardless of how **B** is defined, we can nevertheless conclude that also for multiprobe covariances the PME poses a vast improvement over the standard precision matrix estimator.

5 CONCLUSIONS

It was the starting point of our analysis to find a method for using a priori knowledge about the covariance matrix when estimating the precision matrix from simulations. This requires finding an equivalent of the Kaufman–Hartlap correction when only parts of the covariance are estimated. Using the results of Letac & Massam (2004) we partly solved this task by calculating an expansion of the precision matrix and showing how the leading terms of this expansion can be estimated from simulations. Our method enables the use of pre-existing knowledge on the covariance structure to improve the convergence of the PME and to reduce the noise in its estimation. It also has the advantage that the relative uncertainties of the elements of the PME estimate scale with the number of available simulations N_s as $\sim 1/\sqrt{N_s - 1}$, which is typically much smaller than the uncertainties of the standard precision matrix estimator. The latter also depends on the number of data points N_d and scale as $\sim 1/\sqrt{N_s - N_d - 4}$.

We demonstrated that the PME converges even for drastic deviations between the model covariance and the N -body covariance and we also showed that it provides a much less noisy estimate of the parameter likelihood compared to estimating the precision matrix in the standard way. For a DES weak lensing data vector $N_s \gtrsim 8000$ simulations would be required for the standard estimator to reconstruct the likelihood similarly well as the PME with only $N_s = 200$ – even if the model covariance heavily underestimates Gaussian and non-Gaussian covariance parts. If we assume more realistic deviations between model and N -body covariance, up to 24 000 simulations would be needed for the standard estimator to reconstruct the 1σ quantile of the parameter distribution at the same precision as the PME with only 200 simulations. For an LSST-like weak lensing data vector with $N_d = 2200$ we found that up to 115 000 simulations would be required for the standard estimator to reconstruct the 1σ quantile as well as the PME with only 2400 simulations. It should however be stressed that these statements depend on the quality of the model covariance **M** that was used to compute the PME.

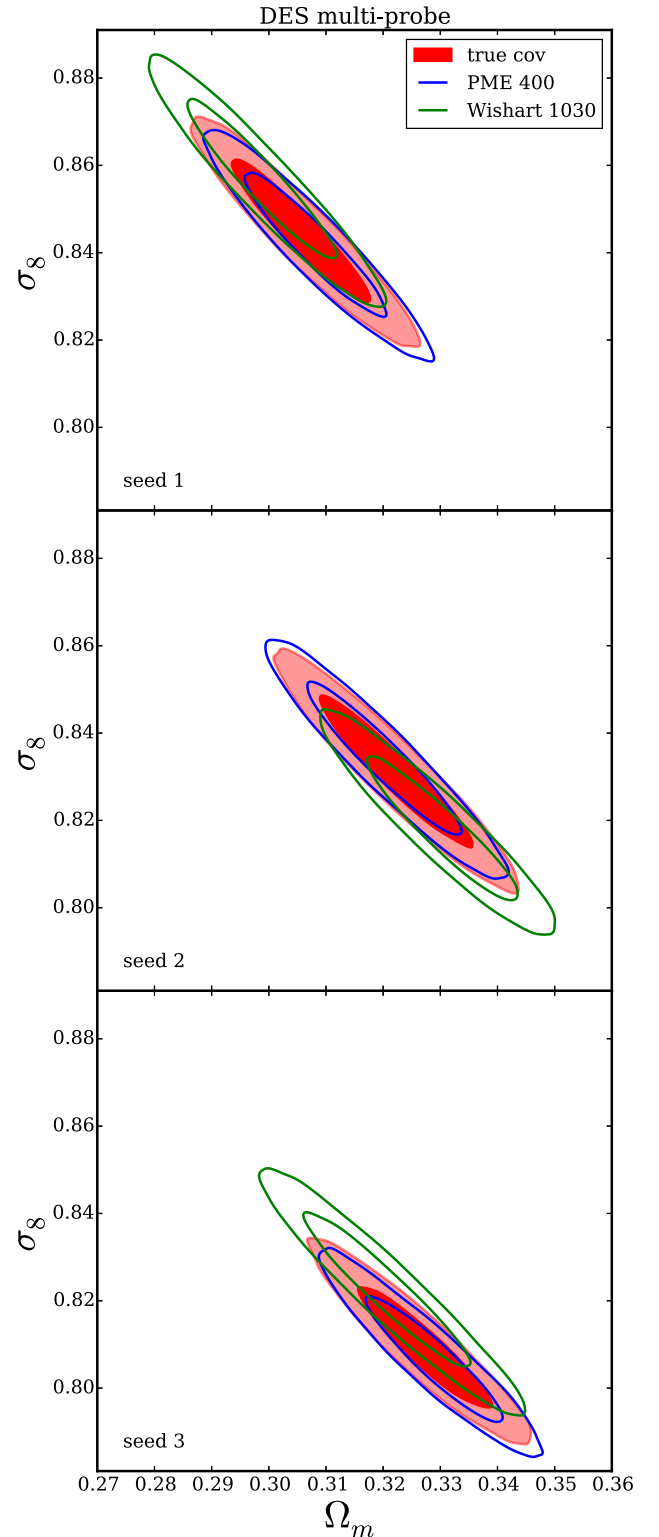


Figure 7. Same as Fig. 2 but for the multiprobe data vector. $N_s = 400$ simulations were assumed for the estimation of the PME while $N_s = N_d + 400 = 1030$ simulations were assumed for the standard estimator. Even with fewer simulations the PME is much better in reconstructing the contours that would be obtained from the true precision matrix of our mock experiment. However, below $N_s = 1600$ a significant offset of the contours persists.

Additional complications arise when galaxy clustering correlation functions are included in the data vector. A performance similar to the weak lensing case can still be achieved if one manages to estimate the cosmic variance of the correlation functions directly, i.e. without shot noise. For this case, we find that a DES-like multiprobe data vector requires up to 44 000 simulations for the standard precision matrix estimator to reconstruct the 1σ quantile of the parameter distribution as well as the PME with 1600 simulations.

One aspect that should be addressed in future work, is to find *a priori* criteria for the convergence of the PME. In Appendix C we demonstrate that it converges for very strong deformations of the halo-model covariance, but one cannot be certain whether and how fast it will converge for all possible data vectors and covariance models. As we show in Appendix C, situations where the PME does not converge can at least be identified *a posteriori* by a comparison of the first-order and second-order expansion. A strong oscillation of likelihood contours derived from the first-order and second-order PME indicates a significant deviation of model and N -body covariance. This way, the PME provides a clear criterion for testing covariance models with simulations – even when the number of available simulations is small.

ACKNOWLEDGEMENTS

This work was supported by SFB-Transregio 33 ‘The Dark Universe’ by the Deutsche Forschungsgemeinschaft (DFG). We also acknowledge the support by the DFG Cluster of Excellence ‘Origin and Structure of the Universe’. The simulations have been carried out on the computing facilities of the Computational Center for Particle and Astrophysics (C2PAP). This research is partially supported by NASA ROSES ATP 16-ATP16-0084 grant. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

This paper has gone through internal review by the DES collaboration. We want to thank Eric Baxter, Gary Bernstein, Scott Dodelson, Franz Elsner, Juan Garcia-Bellido, John Peacock, Stella Seitz and the DES multiprobe pipelines group for very helpful comments and discussions on this project. We also thank the anonymous referee for his or her helpful comments.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The DES data management system is supported by the National Science Foundation under Grant Number AST-1138766.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the Uni-

versity of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex and Texas A&M University.

The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2012-39559, ESP2013-48274, FPA2013-47986, and Centro de Excelencia Severo Ochoa SEV-2012-0234. Research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) including ERC grant agreements 240672, 291329 and 306478.

REFERENCES

- Anderson T., 2003, *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, NJ
- Becker M. R. et al., 2016, *Phys. Rev. D*, 94, 022002
- Dodelson S., Schneider M. D., 2013, *Phys. Rev. D*, 88, 063537
- Eifler T., Schneider P., Hartlap J., 2009, *A&A*, 502, 721
- Eifler T., Krause E., Schneider P., Honscheid K., 2014, *MNRAS*, 440, 1379
- Flaugher B., 2005, *Int. J. Mod. Phys. A*, 20, 3121
- Friedrich O., Seitz S., Eifler T. F., Gruen D., 2016, *MNRAS*, 456, 2662
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
- Heymans C. et al., 2013, *MNRAS*, 432, 2433
- Hilbert S., Hartlap J., Schneider P., 2011, *A&A*, 536, A85
- Ivezic Z. et al., 2009, *Am. Astron. Soc. Meeting Abstr.*, 213, 460
- Joachimi B., 2017, *MNRAS*, 466, L83
- Kaufman G. M., 1967, Report No. 6710. Center for Operations Research and Econometrics, Catholic University of Louvain, Heverlee, Belgium
- Kilbinger M. et al., 2013, *MNRAS*, 430, 2200
- Krause E., Eifler T., 2017, *MNRAS*, 470, 2100
- Letac G., Massam H., 2004, *Scand. J. Stat.*, 31, 295
- Padmanabhan N., White M., Zhou H. H., O’Connell R., 2016, *MNRAS*, 460, 1567
- Paz D. J., Sánchez A. G., 2015, *MNRAS*, 454, 4326
- Pope A. C., Szapudi I., 2008, *MNRAS*, 389, 766
- Rozo E. et al., 2016, *MNRAS*, 461, 1431
- Sellentin E., Heavens A. F., 2016, *MNRAS*, 456, L132
- Sellentin E., Heavens A. F., 2017, *MNRAS*, 464, 4658
- Taylor A., Joachimi B., 2014, *MNRAS*, 442, 2728
- Taylor A., Joachimi B., Kitching T., 2013, *MNRAS*, 432, 1928

APPENDIX A: INFLUENCE OF NOISY COVARIANCE ESTIMATES ON THE SCATTER OF BEST-FITTING COSMOLOGICAL PARAMETERS

Using a noisy precision matrix estimate $\hat{\Psi}$ to determine the best-fitting cosmological parameters by means of

$$\begin{aligned} & \left(\hat{\xi} - \xi[\hat{\pi}_{\text{ML}}] \right)^T \hat{\Psi} \left(\hat{\xi} - \xi[\hat{\pi}_{\text{ML}}] \right) \\ & = \min \left\{ \left(\hat{\xi} - \xi[\pi] \right)^T \hat{\Psi} \left(\hat{\xi} - \xi[\pi] \right) \right\}. \end{aligned} \quad (\text{A1})$$

leads to an additional scatter in these parameters. Especially, this additional noise is not accounted for by the width of contours generated from the precision matrix estimate. This effect has e.g.

been described by Dodelson & Schneider (2013) who also derived a prediction for the additional noise assuming a Gaussian parameter likelihood. They find that the actual parameter covariance when using an inverse-Wishart realization of the precision matrix is given by

$$\mathbf{C}_{\hat{\pi}_{\text{ML}}} = \mathbf{F}^{-1} \left(1 + \frac{(N_d - N_p)(N_s - N_d - 2)}{(N_s - N_d - 1)(N_s - N_d - 4)} \right), \quad (\text{A2})$$

where N_p is the number of considered parameters and \mathbf{F} is the Fisher matrix computed from the true precision matrix of the data vector $\hat{\xi}$. Hence, in the case of a Gaussian parameter likelihood, best-fitting parameters $\hat{\pi}_{\text{ML}}$ that are computed from a Wishart realization of the covariance have also a Gaussian distribution but with a rescaled parameter covariance.

APPENDIX B: UNBIASED ESTIMATOR OF THE SQUARE OF A WISHART MATRIX

Let $\hat{\mathbf{C}}$ be distributed according to a Wishart distribution with ν degrees of freedom and expectation value \mathbf{C} . Then

$$\langle \hat{\mathbf{C}}^2 \rangle \neq \mathbf{C}^2. \quad (\text{B1})$$

However, using the results of Letac & Massam (2004) it is possible to devise an unbiased estimator of \mathbf{C}^2 . It is given by

$$\widehat{(\mathbf{C}^2)} = \frac{\nu^2 \hat{\mathbf{C}}^2 - \nu \hat{\mathbf{C}} \text{tr} \hat{\mathbf{C}}}{\nu^2 + \nu - 2}, \quad (\text{B2})$$

where $\text{tr} \hat{\mathbf{C}}$ denotes the trace of $\hat{\mathbf{C}}$. Using this formula, it is straightforward to derive the estimator of the second-order PME given in equation (12).

APPENDIX C: GENERAL PROPERTIES AND CONVERGENCE OF THE POWER SERIES

C1 General properties

In order to derive some general properties of the PME series, let us slightly change the notation of Section 3. First, let $\mathbf{M}^{1/2}$ be the unique symmetric and positive definite matrix such that

$$\mathbf{M}^{1/2} \mathbf{M}^{1/2} = \mathbf{M}. \quad (\text{C1})$$

This matrix exists as long as our covariance model \mathbf{M} is positive definite. Let us then re-define

$$\mathbf{X} = \mathbf{M}^{-1/2} (\mathbf{B} - \mathbf{B}_m) \mathbf{M}^{-1/2}, \quad (\text{C2})$$

where $\mathbf{M}^{-1/2}$ is the inverse of $\mathbf{M}^{1/2}$, and \mathbf{B} and \mathbf{B}_m are the same as in Section 3. The complete covariance can then be written as

$$\mathbf{C} = \mathbf{M}^{1/2} (\mathbb{1} + \mathbf{X}) \mathbf{M}^{1/2} \quad (\text{C3})$$

and the precision matrix expansion now reads

$$\begin{aligned} \Psi &= \mathbf{M}^{-1/2} \left(\sum_{k=0}^{\infty} (-1)^k \mathbf{X}^k \right) \mathbf{M}^{-1/2} \\ &= \mathbf{M}^{-1/2} (\mathbb{1} - \mathbf{X} + \mathbf{X}^2 + \mathcal{O}[\mathbf{X}^3]) \mathbf{M}^{-1/2}. \end{aligned} \quad (\text{C4})$$

Since both $\mathbf{M}^{-1/2}$ and \mathbf{X} are symmetric matrices, it is immediately clear that this gives a symmetric approximation of Ψ at each order of the power series. The series converges if and only if all eigenvalues of \mathbf{X} fulfill

$$|\lambda_i| < 1, \quad i = 1, \dots, N_d. \quad (\text{C5})$$

In each eigendimension of \mathbf{X} the series $(\mathbb{1} - \mathbf{X} + \mathbf{X}^2 + \mathcal{O}[\mathbf{X}^3])$ is simply the geometric series. For $|\lambda_i| < 1$ the value of this series is >0 at each finite order. At second order, the value of this series is >0 regardless of the values of λ_i . Hence, the second-order PME is always positive definite.

C2 Special cases

C2.1 Rescaling of the covariance

Let us investigate the convergence properties of the power series in equation (10) in a couple of special cases. We start by assuming that our model for the covariance matrix, \mathbf{M} , under- or overestimates the true covariance matrix by a constant factor α , i.e.

$$\mathbf{M} = \alpha \mathbf{C}. \quad (\text{C6})$$

In this case we have

$$\begin{aligned} \mathbf{X} &= \mathbf{M}^{-1/2} (\mathbf{C} - \alpha \mathbf{C}) \mathbf{M}^{-1/2} \\ &= \frac{1 - \alpha}{\alpha} \mathbf{C}^{-1/2} \mathbf{C} \mathbf{C}^{-1/2} \\ &= \frac{1 - \alpha}{\alpha} \mathbb{1}. \end{aligned} \quad (\text{C7})$$

Hence, all eigenvalues of \mathbf{X} are given by $\lambda = \frac{1 - \alpha}{\alpha}$. This has absolute value smaller than 1 for all $\alpha > 0.5$. This especially means that the series used to define the PME converges even if the model covariance overestimates the true covariance by an arbitrarily high overall factor. Since we cut equation (10) after the second order we must however look at how well the series is converged after that order. The relative error on each element of the precision matrix is given by

$$\frac{\Psi_{ij} - \Psi_{2\text{nd},ij}}{\Psi_{ij}} = \lambda^3 = \frac{(1 - \alpha)^3}{\alpha^3}. \quad (\text{C8})$$

This is < 10 per cent for $\alpha \in [0.69, 1.86]$ and < 1 per cent for $\alpha \in [0.83, 1.27]$.

C2.2 Partial rescaling of the covariance

Now let us assume that \mathbf{C} falls into two contributions \mathbf{A} and \mathbf{B} and that only \mathbf{B} is mischaracterized by a constant factor in our model

$$\mathbf{M} = \mathbf{A} + \alpha \mathbf{B}. \quad (\text{C9})$$

Let us furthermore assume that \mathbf{B} has a dominant eigenvalue λ and that \mathbf{v} is an eigenvector to it. If

$$|\lambda \mathbf{v} - \mathbf{C} \mathbf{v}| \ll |\lambda \mathbf{v}| \quad (\text{C10})$$

then the matrix \mathbf{C} and – for values of α that are not too small – also the matrix \mathbf{M} will have an eigendimension close to that of \mathbf{B} with eigenvalues $\lambda_C \approx \lambda$ and $\lambda_M \approx \alpha \lambda$. As a consequence, the matrix \mathbf{X} will have an eigendimension with eigenvalue close to $\lambda_X \approx \frac{1 - \alpha}{\alpha}$ which allows the same conclusion in Section C2.1.

In Section 4 we considered a deformation of the halo model covariance of the form $\mathbf{M} = \mathbf{A} + \mathbf{B}_m$ with

$$\mathbf{B}_m = \alpha \mathbf{C}^{\text{ss,Gauss}} + \beta (\mathbf{C}^{\text{ss,halo}} - \mathbf{C}^{\text{ss,Gauss}}). \quad (\text{C11})$$

This is similar to the situation described above. To illustrate how the rescaling factors α and β impact the convergence of the PME we can e.g. compare the Fisher contours derived from \mathbf{C}^{-1} , \mathbf{M}^{-1} , $\Psi_{1\text{st}}$ and $\Psi_{2\text{nd}}$. In Fig. C1 we show the 1σ and 2σ Fisher-contours for the parameter pair $\Omega_m - \sigma_8$ derived for the DES multiprobe data vector

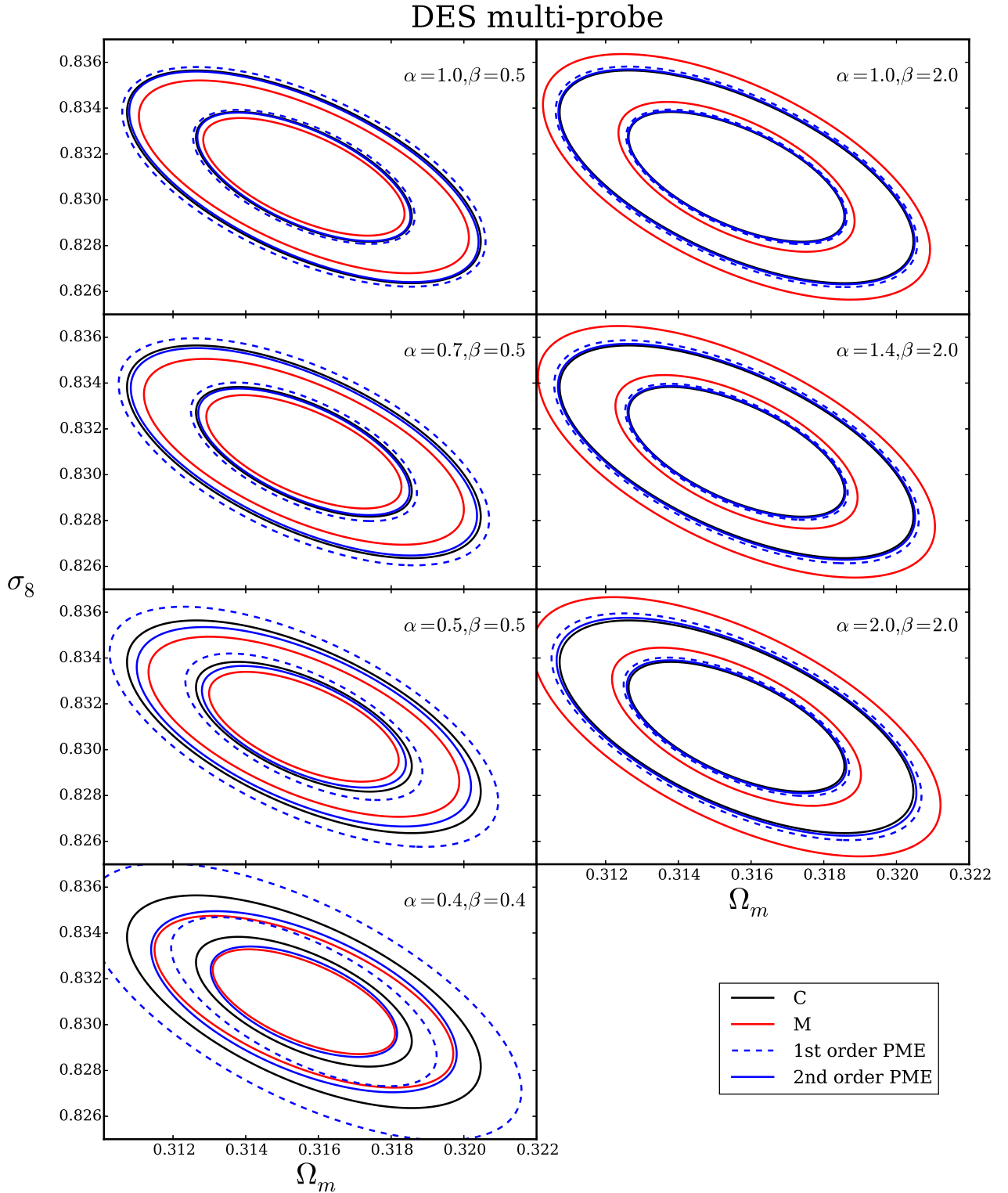


Figure C1. We show the 1σ and 2σ Fisher contours in the Ω_m – σ_8 plane around our fiducial cosmology using the DES multiprobe data vector and keeping all other cosmological parameters fixed. For the black contours the Fisher matrix was derived from the fiducial covariance matrix \mathbf{C} of our experiment – the halo-model covariance. For the red contours we rescaled the Gaussian and non-Gaussian parts of the cosmic variance in \mathbf{C} by constant factors α and β to create our model covariance matrix \mathbf{M} (cf. equation 15). The blue contours show the constraints derived from the first-order PME (dashed lines) and second-order PME (solid lines) of \mathbf{C} around \mathbf{M} . The PME manages to significantly correct the miss estimation of the Fisher matrix by the model precision matrix for most values of the rescaling factors. Only for $\alpha, \beta < 0.5$ the convergence of the PME seems to break down and a strong oscillation between first-order and second-order correction occurs. We discuss this behaviour in detail in Appendix C where we also study examples of more complicated deviations between \mathbf{M} and \mathbf{C} .

using different values of α and β . The figure shows that the PME manages to correct the bias between contours derived from \mathbf{C}^{-1} and contours derived from \mathbf{M}^{-1} even for rather drastic choices of the rescaling factors. Especially for $\alpha, \beta > 1.0$ the convergence is very robust. As predicted by our considerations above, it however breaks down for $\alpha, \beta < 0.5$ where one can see strong oscillations between Ψ_{1st} and Ψ_{2nd} . The convergence of the contours in Fig. C1 is very similar when other parameter combinations are considered or when the contours are derived for the other data vectors considered in this paper.

C2.3 Log-normal motivated approximation to the halo-model covariance

Motivated by the work of Hilbert, Hartlap & Schneider (2011) on approximating the shear–shear covariance matrix with a log-normal approach (cf. their equation 26) we approximate the non-Gaussian parts of the covariance of shear correlation functions as

$$\langle \Delta \xi_{\pm}^A(\theta_i) \Delta \xi_{\pm}^B(\theta_j) \rangle_{\text{non Gauss.}} = \xi_{\pm}^A(\theta_i) \xi_{\pm}^B(\theta_j) R_{AB} \quad (\text{C12})$$

where θ_i labels the different angular bins, A and B label the different auto- and cross-correlation functions and R_{AB} is just a constant factor (depending only on the pair A, B and not on whether ξ_+ or ξ_- are involved). We fix the values of R_{AB} by demanding that our approximation coincides with the halo-model for $\langle \Delta \xi_{+}^A(\theta) \Delta \xi_{+}^B(\theta) \rangle_{\text{non Gauss.}}$ where θ is a certain angular scale which we chose to be either our smallest angular bin ($\theta \approx 3'$) or a slightly larger scale ($\theta \approx 20'$). Note that this is a very crude approximation – even to the log-normal model by Hilbert et al. (2011) since they have not even considered cross-correlations between redshift bins.

We nevertheless use the above matrix as our model covariance \mathbf{M} for the DES shear-shear data vector and compare it to the halo-model covariance \mathbf{C} and the PME. All eigenvalues of matrix \mathbf{X} have in that case $|\lambda_i| < 1$. The three most dominant eigenvalues are

$$\begin{aligned} \lambda_1 &= 0.776 \\ \lambda_2 &= -0.675 \\ \lambda_3 &= 0.197 \end{aligned} \quad (\text{C13})$$

in the case where we match the amplitudes of \mathbf{M} and \mathbf{C} at $\theta \approx 20'$ and

$$\begin{aligned} \lambda_1 &= 0.966 \\ \lambda_2 &= -0.442 \\ \lambda_3 &= 0.203 \end{aligned} \quad (\text{C14})$$

when we match the amplitudes at $\theta \approx 3'$. In both of these cases the PME in principle converges. However, in the second case at least one eigenvalue comes dangerously close to 1. In Fig. C2 we show that in terms of the Fisher contours in the Ω_m – σ_8 plane the PME nevertheless converges and significantly corrects for the deviations between halo-model and log-normal motivated covariance. We have also checked other parameter combinations and find similar results. The reason that a matching at larger scales gives smaller eigenvalues (i.e. better agreement between halo-model and log-normal motivated covariance) is probably that the scaling of equation (C12) fails at small scales.

C2.4 Scale-dependent rescaling of the cosmic variance of the multiprobe data vector

Another alternative way to deform the halo-model covariance is to apply different rescaling factors α and β for the Gaussian and

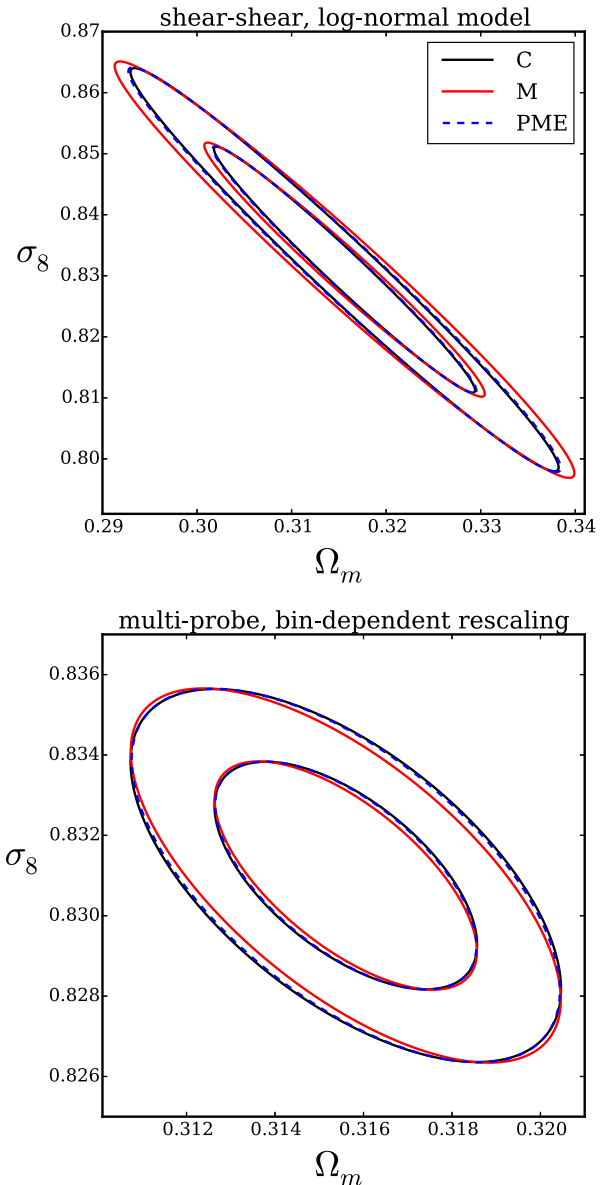


Figure C2. Top: 1σ and 2σ Fisher contours in the Ω_m – σ_8 plane for the DES weak lensing data vector. The black contours are derived from our fiducial halo-model covariance \mathbf{C} . For the red contours we used a model covariance \mathbf{M} that was motivated from the general structure of the log-normal covariance model for shear–shear correlation functions by Hilbert et al. (2011, see main text). The PME (blue dashed contours) still manages to correct for the deviation between the two models. It should however be noted that in this case one eigenvalue of the deviation matrix \mathbf{X} comes dangerously close to one ($\lambda_{\max} = 0.966$). As we discuss in the main text, this situation stabilizes if we match the amplitudes of the halo-model and the log-normal motivated covariance at intermediate angular scales ($\theta \sim 20$ arcmin) instead of the smallest scale of our data vector ($\theta \sim 3$ arcmin). Bottom: we applied a scale-dependent rescaling of the halo-model covariance for the multiprobe data vector motivated by findings of Friedrich et al. (2016). The PME converges also in this case.

non-Gaussian cosmic variance parts for different angular scales (cf. equation 15). If e.g. the finite area of a survey is not correctly accounted for in a covariance model, the results of Friedrich et al. (2016) indicate that this leads to a scale-dependent mischaracterization of the Gaussian cosmic variance and to an almost scale-independent over- or underestimation of the non-Gaussian

parts. Covariance parts involving shape or shot noise on the other hand are less sensitive to the survey area (only to the product of area and galaxy density which is the total number of galaxies).

Motivated by this we replace equation (15) by

$$\mathbf{B}_{m,ij} = \alpha_{ij} \mathbf{C}_{ij}^{\text{ss,Gauss}} + \beta \left(\mathbf{C}_{ij}^{\text{ss,halo}} - \mathbf{C}_{ij}^{\text{ss,Gauss}} \right), \quad (\text{C15})$$

where we choose $\beta = 0.5$ and $\alpha_{ij} = \sqrt{a_i a_j}$ setting a_i to 1.0 at the smallest scales and to 0.5 at the largest scales of the data vector and linearly interpolating for intermediate bins (interpolating in terms of the bin-index).

The most dominant eigenvalues of the matrix \mathbf{X} for this choice of the matrix \mathbf{M} are

$$\begin{aligned} \lambda_1 &= 0.709 \\ \lambda_2 &= -0.440 \\ \lambda_3 &= 0.242, \end{aligned} \quad (\text{C16})$$

i.e. the PME converges. The bottom panel of Fig. C2 also shows that the Fisher contours derived from the second-order PME around this model almost coincide with the ones derived from the halo-model covariance again.

C3 Convergence in the general case

Let us now consider the general case. We want invert the equation

$$\mathbf{C} = \mathbf{M}^{1/2} (\mathbb{1} + \mathbf{X}) \mathbf{M}^{1/2} \quad (\text{C17})$$

where

$$\mathbf{X} := \mathbf{M}^{-1/2} (\mathbf{B} - \mathbf{B}_m) \mathbf{M}^{-1/2}. \quad (\text{C18})$$

Since both \mathbf{M} and \mathbf{C} are positive definite matrices we can immediately infer that also the matrix $\mathbb{1} + \mathbf{X}$ must be positive definite, i.e. all its eigenvalues must be greater than 0. As a consequence, all eigenvalues λ_i of \mathbf{X} must fulfil

$$\lambda_i > -1 \quad \forall i. \quad (\text{C19})$$

In order to invert $\mathbb{1} + \mathbf{X}$ let us change into the eigenbasis of \mathbf{X} by means of an orthogonal matrix \mathbf{U} , i.e.

$$\mathbb{1} + \mathbf{X} = \mathbf{U}^T \text{diag}(1 + \lambda_i) \mathbf{U}. \quad (\text{C20})$$

It is not a priori clear whether we can invert this by means of the geometric series, since we do not know a priori that $|\lambda_i| < 1$. As discussed in Section 5 in the case that $|\lambda_i| > 1$ the PME can at least help to identify differences between a covariance model and covariance from (possibly very few) simulations since in that case the first-order and second-order PME will display a divergent behaviour. However, since we know a priori that $\lambda_i > -1$ we can in principle apply a trick to let the PME series converge in any case. This trick is to expand $1/(1 + \lambda)$ not around $\lambda_0 = 0$ but around some other point $\lambda_0 = a > 0$:

$$\frac{1}{1 + \lambda} = \frac{1}{1 + a} \left[1 - \left(\frac{x - a}{1 + a} \right) + \left(\frac{x - a}{1 + a} \right)^2 - \dots \right]. \quad (\text{C21})$$

In terms of the PME series this is in fact equivalent to replacing the model covariance \mathbf{M} by $(1 + a)\mathbf{M}$. This way, one can in principle

always ensure convergence of the series. This however comes at the expense of the series converging very slowly for eigenvalues of \mathbf{X} that are already close to or smaller than 0. Since in a real case scenario \mathbf{M} is assumed to be our best guess for the true covariance we hence recommend to stay with $a = 0$ and interpret a divergent PME as a significant difference between model and N -body covariance.

APPENDIX D: DATA VECTORS

D1 Weak lensing data vectors

The redshift distribution and tomographic binning used for our LSST-like weak lensing data vector was chosen to be exactly that of Krause & Eifler (2017, see section 3). This means we assumed an overall source density of 26 arcmin^{-2} and a source distribution with a median redshift of ≈ 0.7 that extends out to $z \gtrsim 3.0$. The tomographic bins were defined by first splitting the redshift distribution into 10 non-overlapping bins of equal source density and then assuming a Gaussian photoz uncertainty of $\sigma_z = 0.05$. The intrinsic ellipticity dispersion of the sources was assumed to be $\sigma_\epsilon = 0.26$ per ellipticity component.

The redshift distribution for the DES-like data vector was chosen to be shallower as for the LSST case reflecting the smaller depth of DES. Here our source distribution has a median redshift of ≈ 0.5 and extends out to $z = 2.0$. The overall source density was taken to be $10/\text{arcmin}^2$ and the five tomographic bins were defined assuming a photoz uncertainty of $\sigma_z = 0.08$. The intrinsic ellipticity dispersion was chosen to be the same as for the LSST-like case.

D2 Lens galaxies

For the DES multiprobe data vector we also considered galaxy clustering and galaxy–galaxy lensing correlation functions. For this we were assuming a sample of foreground galaxies with a constant comoving density motivated by the DES redMaGiC sample (Rozo et al. 2016) divided into three tomographic bins whose redshift ranges are (0.20, 0.35), (0.35, 0.50) and (0.50, 0.65). For these galaxies we assumed zero redshift uncertainties motivated by the fact that the redMaGiC redshift errors are small compared to the values for our source samples. The overall density of foreground galaxies was taken to be $0.15/\text{arcmin}^2$.

D3 Binning and scales

The real space data vectors use 15 logarithmic angular bin from $\theta = 2.5'$ to $\theta = 250'$ for each correlation function and the Fourier space data vector uses 40 logarithmic bin from $\ell = 20$ to $\ell = 5000$ for each power spectrum. Data vector I contains the correlation functions ξ_+ and ξ_- for each possible combination of source bins. Data vector Ia also contains the autocorrelation of the lens bins and all possible combinations of lens–source correlations (i.e. only those combinations where the sources are at higher redshifts than the lenses). Data vector II contains the auto- and cross-power spectra of all possible combinations of source bins.

This paper has been typeset from a \LaTeX file prepared by the author.